

A logical analysis of Monty Hall and Sleeping Beauty

Allen L. Mann* and Ville Aarnio

January 12, 2017

*The first author wishes to gratefully acknowledge the partial support of the European Science Foundation EU-ROCORES program LogICCC [FP002—Logic for Interaction (LINT)] and the Academy of Finland (grant 129208).

Abstract

Hintikka and Sandu's independence-friendly (IF) logic is a conservative extension of first-order logic that allows one to consider semantic games with imperfect information. In the present article, we first show how several variants of the Monty Hall problem can be modeled as semantic games for IF sentences. In the process, we extend IF logic to include semantic games with chance moves and dub this extension *stochastic IF logic*. Finally, we use stochastic IF logic to analyze the Sleeping Beauty problem, leading to the conclusion that the thirders are correct while identifying the main error in the halfers' argument.

1 Introduction

A game-show contestant is presented with three doors, one of which conceals a prize. After the contestant selects a door, the host (Monty Hall) opens one of the two remaining doors, being careful not to reveal the prize. He then offers the contestant the opportunity to switch doors. Should the contestant stick with her original choice or switch?¹

The surprising answer is that the contestant should always switch doors. If she does, she will win with probability $2/3$, whereas if she sticks with her original door she will win with probability $1/3$. What is even more surprising is that the Monty Hall problem can be expressed by the independence-friendly sentence φ_{MH} ,

$$\forall x(\exists y/\{x\})\forall z[(z \neq x \wedge z \neq y) \implies (\exists y/\{x\})x = y],$$

interpreted in the structure $\mathbf{M} = \{1, 2, 3\}$. The sentence can be read:

For any Door x concealing the prize, there exists a Door y chosen (independently of x) by the contestant such that for every possible Door z opened by Monty Hall that differs from the door containing the prize and the door chosen by the contestant, there is a Door y chosen (independently of x) by the contestant such that $x = y$.

To understand the connection between the Monty Hall problem and φ_{MH} , one must interpret the formula game theoretically. The semantic game for a first-order sentence is a contest between two players. The existential player attempts to verify the sentence by choosing the values of existentially quantified variables, while the universal player tries to falsify the sentence by picking the values of universally quantified variables. Disjunctions prompt the verifier to choose a disjunct; conjunctions prompt the falsifier to pick a conjunct. Negation tells the players to switch roles. The sentence is true if the existential player has a winning strategy, false if the universal player has a winning strategy. Traditionally, the existential player is named Eloise and universal player Abelard.

The semantic game for a first-order sentence is a game with perfect information in the sense that, at each decision point, the active player is aware of all prior moves. In contrast, the semantic game for φ_{MH} is a game with imperfect information since Eloise must choose the value of y without knowing the value of x . Thus, the sentence φ_{MH} is “independence-friendly” in the sense that the contestant’s choices may not depend on the location of the prize.

The present paper is organized as follows. The next section presents the Monty Hall problem as an extensive game, while Section 3 introduces the syntax and semantics of independence-friendly (IF) logic. We then show that the Monty Hall problem is equivalent to the semantic game for φ_{MH} and consider a variant of the problem in which the host is not required to offer the contestant the opportunity to switch doors. In Section 4, we consider another variant of the Monty Hall problem in which the host is indifferent to the outcome of the game. To analyze this variant, we extend IF logic by adding a third player (Nature) who makes moves at random. We then use this extension in Section 5 to address the controversy surrounding the Sleeping Beauty problem, which has divided the philosophical community into two camps, thirders and halvers, who respectively defend the contradictory solutions $1/3$ and $1/2$. We briefly review the arguments presented by both sides, and present several ways to formalize the problem using stochastic IF logic. Under our preferred formalization the answer is $1/3$, but under two alternate formalizations the answer is $1/2$.

¹The Monty Hall problem was popularized by Marilyn vos Savant [26]. For an overview of the history of the problem and its wider implications, see Tierney [22, 23].

2 The Monty Hall problem as an extensive game

In an extensive game, players take turns making moves until the game ends, at which point each player receives a certain payoff. The following more formal definition is taken with slight modifications from Osborne and Rubinstein [19, p. 200].

Definition. An *extensive game with imperfect information* has the following components.

- A finite set N of *players*.
- A set H of sequences (called *histories*) closed under initial segments.
 - A component a_i of a history $h = (a_1, \dots, a_n)$ is called an *action*.
 - A history of the form $h \frown a = (a_1, \dots, a_n, a)$ is called a *successor* of h .
 - A *terminal history* is a history with no successors. The set of terminal histories is denoted $Z \subseteq H$; the set of actions available after the nonterminal history $h \in H \setminus Z$ is denoted

$$A(h) = \{a : h \frown a \in H\}.$$

- A *player function* $P: H \setminus Z \rightarrow N$ that indicates whose turn it is to move. Let $H_p = P^{-1}(p)$ denote the set of histories after which it is player p 's turn.
- For each player $p \in N$, an equivalence relation \sim_p on H_p with the property that for all $h, h' \in H_p$,

$$h \sim_p h' \text{ implies } A(h) = A(h').$$

If $h \sim_p h'$, we say the histories h and h' are *indistinguishable* to player p . An equivalence class

$$[h]_{\sim_p} = \{h' \in H_p : h \sim_p h'\}$$

is called an *information set* for player p .

- A *utility function* $u: Z \rightarrow \mathbb{R}^N$ that specifies the payoff each player receives at the end of the game. For each terminal history $h \in Z$ and player $p \in N$, let $u_p(h) = u(h)(p)$ denote the payoff received by player p . A *constant-sum* game is one in which the sum of the payoffs received by the players always takes the same value. In a *win-lose* game, a single player (the *winner*) receives a payoff of 1, while all other players receive a payoff of 0.

An extensive game can also be represented by a game tree [15, page 42]. The game tree for the Monty Hall problem is shown in Figure 1, where for later convenience we will treat the contestant as player I, and Monty Hall as player II. First, Monty Hall secretly places the grand prize behind one of three doors. Second, the contestant guesses which door conceals the prize. Third, Monty opens a door, revealing its contents. He never opens the door initially chosen by the contestant, however, nor does he reveal the prize. Thus, if the contestant guessed correctly, Monty Hall is free to open either of the two remaining doors. If the contestant guessed incorrectly, however, Monty is forced to open the only other door that does not conceal the prize. Finally, Monty Hall offers the contestant the opportunity to stick with her original guess or to switch to the other unopened door. The contestant wins if she chooses the door concealing the grand prize; otherwise Monty Hall wins.

What makes the Monty Hall problem interesting is the contestant's uncertainty about which door conceals the prize. Thus, she is unsure which node of the game tree corresponds to the current

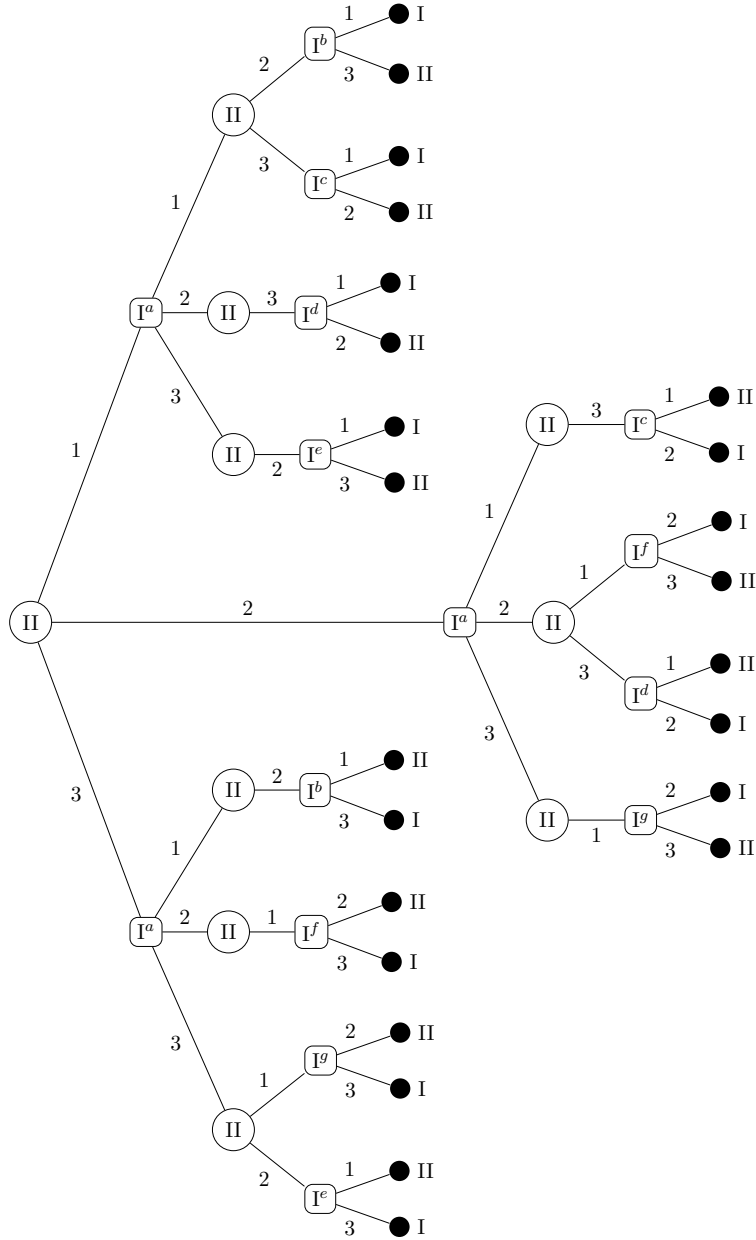


Figure 1: The Monty Hall problem

state of the game. In Figure 1, we label positions that are indistinguishable to the contestant with the same superscript letter. For example, the contestant cannot distinguish between the positions labeled I^a because she does not know the location of the prize when making her initial choice. Similarly, in both positions labeled I^b , the contestant initially chose Door 1, after which Monty Hall opened Door 2.

Notice that the game tree shows which actions Monty Hall and the contestant may take, but not which actions they *should* take. For that, each player needs a strategy.

Definition. A *pure strategy* is a rule that tells a player how to move whenever it is his or her turn. In other words, a pure strategy for player p is a choice function

$$\sigma \in \prod_{h \in H_p} A(h)$$

that respects the player's indistinguishability relation \sim_p , that is,

$$\sigma(h) = \sigma(h') \quad \text{whenever} \quad h \sim_p h'.$$

The set of pure strategies for player p is denoted S_p .

A player p is said to *follow* a strategy $\sigma \in S_p$ in a history $h' \in H$ if, for every history $h \in H_p$ that is a proper initial segment of h' , the history $h \frown \sigma(h)$ is also an initial segment of h' . In a win-lose game, a pure strategy is *winning* if its player wins every terminal history in which he or she follows it.

The outcome of an extensive game is determined once every player has chosen a strategy. For example, suppose Monty Hall decides to place the prize behind Door 1 and to open the lowest-numbered door possible (i.e., not containing the prize or selected by the contestant). For her part, if the contestant selects Door 1, then switches to whichever door is not opened by Monty Hall, she will lose by switching to Door 3 after Monty Hall opens Door 2.

Definition. A *pure-strategy profile* is a vector $\sigma = \langle \sigma_p \in S_p : p \in N \rangle$ of pure strategies, one for each player.

Definition. Let $S = \prod_{p \in N} S_p$ be the set of all pure-strategy profiles, and let $h_\sigma \in Z$ denote the terminal history induced by the pure-strategy profile $\sigma \in S$. By abuse of notation, we will let $u(\sigma) = u(h_\sigma)$ denote the payoff vector received by the players when they follow the pure-strategy profile σ .

Although the contestant in the Monty Hall problem does not have a winning pure strategy, she does have an effective counterstrategy to each of Monty Hall's strategies. For example, if Monty Hall always places the prize behind Door 1, then the contestant can win by choosing Door 1 and sticking with it when offered the opportunity to switch. Similarly, if the contestant always chooses Door 1 and sticks with it, Monty Hall can counter by placing the prize behind Door 2 or Door 3. Thus, the losing player can always improve his or her payoff by changing strategies.

Definition. Let $\sigma = \langle \sigma_p : p \in N \rangle$ be a pure-strategy profile. A *unilateral deviation* from σ by player p is a pure-strategy profile $\sigma' = \langle \sigma'_p : p \in N \rangle$ in which every player other than p follows the same pure strategy as in σ , i.e., for all $q \in N \setminus \{p\}$, we have $\sigma_q = \sigma'_q$. We will follow the standard convention of writing $\sigma' = \langle \sigma'_p, \sigma_{-p} \rangle$. Such a deviation is *profitable* for player p if $u_p(\sigma) < u_p(\sigma')$.

Definition. A *pure-strategy equilibrium* $\sigma^* \in S$ is a pure-strategy profile from which no player has a profitable deviation.

As we observed above, the Monty Hall problem does not have a pure-strategy equilibrium since there are profitable deviations from every strategy profile. However, if we allow the players to vary their pure strategies from one play to the next, we can study the long-run effect of selecting each pure strategy with a given probability.

Definition. A *mixed strategy* for player p is a probability distribution over S_p . The set of mixed strategies for player p is denoted $\Delta(S_p)$.

When the players follow mixed strategies instead of pure strategies, the outcome of the game is no longer determined, but each outcome will occur with a certain probability.

Definition. A *mixed-strategy profile* is a vector $\boldsymbol{\mu} = \langle \mu_p \in \Delta(S_p) : p \in N \rangle$ of mixed strategies, one for each player. If we assume that the players select their pure strategies independently, then each mixed-strategy profile induces a probability distribution μ on $S = \prod_{p \in N} S_p$ defined by

$$\mu(\boldsymbol{\sigma}) = \prod_{p \in N} \mu_p(\sigma_p),$$

where $\boldsymbol{\sigma} = \langle \sigma_p \in S_p : p \in N \rangle$. When S is finite, we can define the *expected utility* of $\boldsymbol{\mu}$ by

$$U(\boldsymbol{\mu}) = \sum_{\boldsymbol{\sigma} \in S} \mu(\boldsymbol{\sigma}) u(\boldsymbol{\sigma}),$$

where the *expected utility for player p* is $U_p(\boldsymbol{\mu}) = U(\boldsymbol{\mu})(p)$.

Example. Consider a two-player, win-lose game in which the first player has three pure strategies, $S_I = \{\sigma_1, \sigma_2, \sigma_3\}$, and the second player has only two pure strategies, $S_{II} = \{\tau_1, \tau_2\}$. Suppose the first player follows a mixed strategy μ such that $\mu(\sigma_1) = \mu(\sigma_2) = \mu(\sigma_3) = 1/3$, while the second player follows a mixed strategy ν such that $\nu(\tau_1) = 2/5$, and $\nu(\tau_2) = 3/5$. Figure 2 depicts the situation where

$$\begin{aligned} u_I(\sigma_1, \tau_1) &= u_I(\sigma_2, \tau_2) = u_I(\sigma_3, \tau_1) = 1, \\ u_I(\sigma_1, \tau_2) &= u_I(\sigma_2, \tau_1) = u_I(\sigma_3, \tau_2) = 0. \end{aligned}$$

The area of the shaded region is equal to $U_I(\mu, \nu) = 7/15$, while the area of the unshaded region is $U_{II}(\mu, \nu) = 8/15$.

One can easily see that neither player's mixed strategy is optimal. If player II follows ν , then player I can improve her chances of winning by following σ_2 more often and σ_1 and σ_3 less often. Conversely, if player I follows μ , then player II can improve his expected utility by following τ_2 more often and τ_1 less often.

The previous example shows that players can profitably deviate from mixed-strategies as well as pure-strategies.

Definition. Let $\boldsymbol{\mu} = \langle \mu_p \in \Delta(S_p) : p \in N \rangle$ be a mixed-strategy profile. A *unilateral deviation* from $\boldsymbol{\mu}$ by player p is a mixed-strategy profile $\boldsymbol{\mu}' = \langle \mu'_p \in \Delta(S_p) : p \in N \rangle$ in which every player other than p follows the same mixed strategy as in $\boldsymbol{\mu}$, i.e., for all $q \in N \setminus \{p\}$, we have $\mu_q = \mu'_q$. We will follow the standard convention of writing $\boldsymbol{\mu}' = \langle \mu'_p, \mu_{-p} \rangle$. Such a deviation is *profitable* for player p if $U_p(\boldsymbol{\mu}) < U_p(\boldsymbol{\mu}')$.

Definition. A *mixed-strategy equilibrium* is a mixed-strategy profile $\boldsymbol{\mu} \in \prod_{p \in N} \Delta(S_p)$ from which no player has a profitable deviation.

	τ_1	τ_2
σ_1		
σ_2		
σ_3		

Figure 2: A pair of mixed strategies

John Nash proved that a strategic game has a mixed-strategy equilibrium if there are finitely many players that each have a finite number of pure strategies [16, 17]. Nash's theorem is a generalization of von Neumann's minimax theorem, which states that every two-player, constant-sum game in which each player has a finite number of pure strategies has a mixed-strategy profile that simultaneously maximizes the minimum utility expected by each player. Player I's expected utility from such a mixed-strategy equilibrium is called the minimax value of the game [25].

We now verify that the minimax value of the Monty Hall problem is $2/3$. Observe that Monty Hall has a total of $3 \cdot 2^3 = 24$ pure strategies. However, pure strategies that differ only at decision points that are never reached when those strategies are followed are *outcome equivalent* [19, page 94]. By identifying outcome-equivalent strategies, we can reduce the number of Monty Hall's strategies by a factor of four. Let τ_a denote the *reduced strategy* [19, page 94] according to which Monty Hall places the prize behind Door a , then opens the door with the lowest number possible, and let τ^a be the reduced strategy according to which he places the prize behind Door a , then opens the door with the highest number possible. Let ν^* be the mixed strategy according to which $\nu^*(\tau_a) = 1/6 = \nu^*(\tau^a)$.

The contestant has $3 \cdot 2^6 = 192$ pure strategies, but only twelve reduced strategies. Let σ_b denote the reduced strategy according to which the contestant initially chooses Door b , then sticks with her initial choice when she is offered the opportunity to switch doors. Let σ'_b denote the reduced strategy according to which the contestant initially chooses Door b , then switches doors. Observe that the contestant has six additional reduced strategies (that will remain nameless) according to which her decision to stick or switch doors depends on the door opened by Monty Hall. Finally, let μ^* denote the mixed strategy according to which $\mu^*(\sigma'_b) = 1/3$.

If the contestant sticks with her original door, she will win whenever it contains the prize, while, if she switches doors, she will win whenever her original door does not contain the prize. This information is summarized in Figure 3, from which it is straightforward to calculate her expected utility $U_I(\mu^*, \nu^*) = 2/3$. To show that neither player has a profitable deviation in mixed strategies, it suffices to fix μ^* and compute how well it fares against each of Monty Hall's reduced strategies, then fix ν^* and compute how it fares against each of the contestant's reduced strategies [15, page 151]. Again, inspecting Figure 3 reveals that

$$\min_{\tau \in S_{\text{IH}}} U_I(\mu^*, \tau) = \frac{2}{3} = \max_{\sigma \in S_I} U_I(\sigma, \nu^*).$$

(Note that for each of the reduced strategies σ not listed in Figure 3 we have $U_I(\sigma, \nu^*) = 1/2$.)

Thus $\langle \mu^*, \nu^* \rangle$ is a mixed-strategy equilibrium, and the contestant's minimax value is $2/3$.²

	τ_1	τ_2	τ_3	τ^1	τ^2	τ^3
σ_1	1	0	0	1	0	0
σ_2	0	1	0	0	1	0
σ_3	0	0	1	0	0	1
σ'_1	0	1	1	0	1	1
σ'_2	1	0	1	1	0	1
σ'_3	1	1	0	1	1	0

Figure 3: Half of the reduced strategic form of the Monty Hall problem. The six rows that are not shown each contain three 0's and three 1's.

3 First-order logic with imperfect information

In this section, we introduce the syntax and game-theoretic semantics of first-order logic with imperfect information. There are several variants found in the literature,³ but we adopt the original slashed notation of independence-friendly (IF) logic used by Hintikka and Sandu [9, 10]. Our presentation will necessarily be brief. For a fuller treatment, we refer the reader to [14].

First-order logic with imperfect information is a conservative extension of first-order logic that includes formulas whose semantic games are extensive games with imperfect information. Ordinary first-order formulas are built up from atomic formulas using negation (\neg), disjunction (\vee), conjunction (\wedge), existential quantification (\exists), and universal quantification (\forall). Atomic formulas and negated atomic formulas are called *literals*. Independence-friendly formulas are similar to first-order formulas except that each connective and quantifier is parameterized by a finite set of variables that specifies the information available to the relevant player.

Definition. Given any first-order vocabulary, an *independence-friendly formula* is an element of the smallest set IF satisfying the following conditions:

- Every first-order literal belongs to IF.
- If $\varphi, \psi \in \text{IF}$, and W is a finite set of variables, then $(\varphi \vee_{/W} \psi)$ and $(\varphi \wedge_{/W} \psi)$ belong to IF.
- If $\varphi \in \text{IF}$, x is a variable, and W is a finite set of variables, then $(\exists x_{/W})\varphi$ and $(\forall x_{/W})\varphi$ belong to IF.

The finite set of variables W is called an *independence set* or *slash set*.

Informally, a slash set indicates the variables in which a player's choice must be uniform. For example, in the semantic game for $\varphi \vee_{/x} \psi$, Eloise's choice of disjunct may not depend on x . In the semantic game for $(\forall z_{/x,y})\varphi$, Abelard's choice of z must be uniform in x and y . When

²A similar analysis of the Monty Hall problem as an extensive game appears in Sandu [20, pages 239–244].

³In particular, Väänänen's dependence logic [24] has attracted a significant following. In dependence logic, the dependence relation between quantified variables is specified by *dependence atoms* such as

$$=(x_1, \dots, x_n, y),$$

which indicates that the value of y is determined by the values of x_1, \dots, x_n .

a slash set is empty we simply omit it. Free and bound variables are defined as usual, with the proviso that variables in slash sets are free.⁴ An *independence-friendly sentence* is an IF formula with no free variables. The set of subformulas of an IF formula φ is denoted $\text{Subf}(\varphi)$; the set of literal subformulas of φ is denoted $\text{Lit}(\varphi)$.

To avoid unnecessary complications, we will assume that all IF formulas are in negation normal form, i.e., the negation symbol only appears in front of atomic formulas. However, for any IF formula φ , we will use the notation $\neg\varphi$ as a recursively defined abbreviation:

$$\begin{aligned}\neg\neg\varphi & \text{ is } \varphi, \\ \neg(\varphi \vee_W \psi) & \text{ is } \neg\varphi \wedge_W \neg\psi, \\ \neg(\varphi \wedge_W \psi) & \text{ is } \neg\varphi \vee_W \neg\psi, \\ \neg(\exists x/W)\varphi & \text{ is } (\forall x/W)\neg\varphi, \\ \neg(\forall x/W)\varphi & \text{ is } (\exists x/W)\neg\varphi.\end{aligned}$$

We will also use $\varphi \implies \psi$ as an abbreviation for $\neg\varphi \vee \psi$.

Now that we have defined the syntax of IF logic, we next present its game-theoretic semantics.

Definition. Let φ be an IF sentence, and let \mathbf{M} be a suitable structure.⁵ The *semantic game* $G(\mathbf{M}, \varphi)$ is defined as follows:

- There are two players, Eloise (\exists) and Abelard (\forall).
- The set of histories is $H = \bigcup \{ H_\psi : \psi \in \text{Subf}(\varphi) \}$, where H_ψ is defined recursively:

- $H_\varphi = \{(\varphi)\}$.
- If ψ is $\chi_1 \vee_W \chi_2$, then $H_{\chi_i} = \{ h \frown \chi_i : h \in H_{\chi_1 \vee_W \chi_2} \}$.
- If ψ is $\chi_1 \wedge_W \chi_2$, then $H_{\chi_i} = \{ h \frown \chi_i : h \in H_{\chi_1 \wedge_W \chi_2} \}$.
- If ψ is $(\exists x/W)\chi$, then $H_\chi = \{ h \frown (x, a) : h \in H_{(\exists x/W)\chi}, a \in M \}$.
- If ψ is $(\forall x/W)\chi$, then $H_\chi = \{ h \frown (x, a) : h \in H_{(\forall x/W)\chi}, a \in M \}$.

Every history h induces an assignment s_h defined by

$$s_h = \begin{cases} \emptyset & \text{if } h = (\varphi), \\ s_{h'} & \text{if } h = h' \frown \psi, \\ s_{h'}(x/a) & \text{if } h = h' \frown (x, a), \end{cases}$$

where $s_{h'}(x/a)$ is the assignment that is identical to $s_{h'}$ except that it assigns the value a to the variable x . For example, let $R(x, y)$ be an atomic formula, and suppose φ is

$$\forall x (\exists y / \{x\}) [R(x, y) \wedge \neg R(x, y)].$$

For any $a, b \in M$, the sequence $h = (\varphi, (x, a), (y, b), \neg R(x, y))$ is a history for $G(\mathbf{M}, \varphi)$ that induces the assignment defined by

$$s_h(x) = a \quad \text{and} \quad s_h(y) = b.$$

⁴For example, the variable x is free in the formula $(\exists y / \{x\})\varphi$, while both x and y are free in $(\exists y / \{x, y\})\varphi$.

⁵A structure is *suitable* for an IF formula φ if it interprets every function, relation, and constant symbol appearing in φ .

- Once play reaches a literal the game ends, i.e., the set of terminal histories is:

$$Z = \bigcup_{\chi \in \text{Lit}(\varphi)} H_\chi.$$

Observe that the above history h is terminal because $\neg R(x, y)$ is a literal.

- Disjunctions and existential quantifiers are decision points for Eloise, while conjunctions and universal quantifiers are decision points for Abelard:

$$P(h) = \begin{cases} \exists & \text{if } h \in H_{\chi_1 \vee_W \chi_2} \text{ or } h \in H_{(\exists x/W)\chi}, \\ \forall & \text{if } h \in H_{\chi_1 \wedge_W \chi_2} \text{ or } h \in H_{(\forall x/W)\chi}. \end{cases}$$

Let $H_\exists = P^{-1}(\exists)$ and $H_\forall = P^{-1}(\forall)$ be the sets of histories in which Eloise and Abelard are respectively active.

- The indistinguishability relations \sim_\exists and \sim_\forall are defined as follows.

- For $h, h' \in H_{\chi_1 \vee_W \chi_2}$ or $h, h' \in H_{(\exists x/W)\chi}$, we have $h \sim_\exists h'$ if and only if

$$\{x \in \text{dom}(s_h) : s_h(x) \neq s_{h'}(x)\} \subseteq W.$$

- For $h, h' \in H_{\chi_1 \wedge_W \chi_2}$ or $h, h' \in H_{(\forall x/W)\chi}$, we have $h \sim_\forall h'$ if and only if

$$\{x \in \text{dom}(s_h) : s_h(x) \neq s_{h'}(x)\} \subseteq W.$$

Returning to our example φ above, if $a, a' \in M$, then the histories $(\varphi, (x, a)) \sim_\exists (\varphi, (x, a'))$ are indistinguishable to Eloise.

- Eloise wins a terminal history $h \in H_\chi$ if $\mathbf{M}, s_h \models \chi$; Abelard wins if $\mathbf{M}, s_h \not\models \chi$.

For example, Eloise wins the above history h if and only if $(a, b) \notin R^\mathbf{M}$.

As with first-order logic, the truth or falsity of an IF sentence is determined not by winning a single play of the game, but by having a winning strategy.

Definition. Let φ be an IF formula, and let \mathbf{M} be a suitable structure. Then φ is *true* in \mathbf{M} , denoted $\mathbf{M} \models^+ \varphi$, if Eloise has a winning strategy for $G(\mathbf{M}, \varphi)$, and it is *false* in \mathbf{M} , denoted $\mathbf{M} \models^- \varphi$, if Abelard has a winning strategy.

The semantic game for a first-order sentence is a two-player, win-lose game with perfect information and finite horizon. Thus, the principle of bivalence for first-order logic is a consequence of the Gale–Stewart theorem [7]. Since the semantic game for an IF sentence is a game with imperfect information, the Gale–Stewart theorem does not apply. Hence, it is possible to have IF sentences that are neither true nor false in a given structure.

Example. Consider the semantic game for the first-order sentence $\forall x \exists y (x = y)$. When played on the two-element structure $\mathbf{2} = \{0, 1\}$, Abelard has two possible strategies, $x := 0$ and $x := 1$, while Eloise has four possible strategies, $y := 0$, $y := 1$, $y := x$, and $y := 1 - x$. Here, $x := 0$ denotes the pure strategy that assigns the value 0 to x , while $y := x$ denotes the pure strategy that assigns y the same value as x .

In contrast, Eloise cannot distinguish the histories $(\varphi, (x, 0)) \sim_\exists (\varphi, (x, 1))$ in the semantic game for the IF sentence $\forall x (\exists y / \{x\}) x = y$. Thus she only has two strategies, $y := 0$ and $y := 1$, neither of which is winning. It is easy to see that neither of Abelard's strategies, $x := 0$ and $x := 1$, are winning either (see Figure 4).

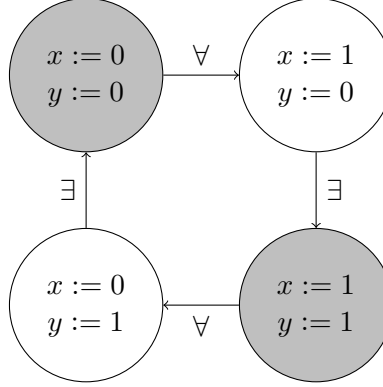


Figure 4: The strategic form of the semantic game for $\forall x(\exists y/\{x\}) x = y$ when played on the structure $\mathbf{2} = \{0, 1\}$. The arrows represent profitable deviations for the indicated player.

Notice that the strategic form of the semantic game in the previous example is equivalent to the game Matching Pennies, in which two players simultaneously turn their respective coins to Heads or Tails. The first player wins if the coins match; the second player wins if they differ. Although Matching Pennies does not have a pure-strategy equilibrium, it does have a mixed-strategy equilibrium where both players turn their coins to Heads or Tails with equal probability. The semantic game for $\forall x(\exists y/\{x\}) x = y$ played on the structure $\mathbf{2} = \{0, 1\}$ has a similar mixed-strategy equilibrium $\langle \mu^*, \nu^* \rangle$, where

$$\begin{aligned}\nu^*(x := 0) &= 1/2 = \nu^*(x := 1), \\ \mu^*(y := 0) &= 1/2 = \mu^*(y := 1),\end{aligned}$$

and the minimax value for Eloise is $U_{\exists}(\mu^*, \nu^*) = 1/2$. In a structure with n elements, the minimax value for Eloise is $1/n$.⁶

Definition. The *truth value* of an IF sentence φ in a suitable finite structure \mathbf{M} is the minimax value for Eloise in the semantic game $G(\mathbf{M}, \varphi)$.

The above definition is due to Sevenster and Sandu [21], who dub their extension of the basic game-theoretic semantics for IF logic *equilibrium semantics*. Galliani [8] independently developed a similar semantics based on behavioral strategies.⁷

We are now ready to show that, when interpreted in the structure $\mathbf{M} = \{1, 2, 3\}$, the semantic game for the sentence φ_{MH} ,

$$\forall x(\exists y/\{x\})\forall z\left[(z \neq x \wedge z \neq y) \implies (\exists y/\{x\}) x = y\right],$$

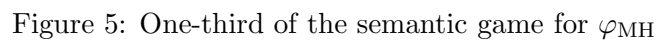
is equivalent to the Monty Hall problem. Note that the above sentence is an abbreviation for

$$\forall x(\exists y/\{x\})\forall z\left[(z = x \vee z = y) \vee (\exists y/\{x\}) x = y\right].$$

At first glance, the semantic game $G(\mathbf{M}, \varphi_{\text{MH}})$ appears more complicated than the Monty Hall problem since the contestant only makes two moves, whereas Eloise makes three moves. Moreover, Abelard and Eloise each have more possible actions in $G(\mathbf{M}, \varphi_{\text{MH}})$ than Monty Hall and the contestant, respectively. Figure 5 shows the part of $G(\mathbf{M}, \varphi_{\text{MH}})$ that occurs after Abelard sets the value of x to 1, which corresponds to the upper third of Figure 1.

⁶This example, due to Ajtai, first appeared in [2].

⁷A *behavioral strategy* for player p is a function mapping each of his or her information sets to a probability distribution over the set of possible actions at that information set.



Abelard's possible strategies in the game $G(\mathbf{M}, \varphi_{\text{MH}})$ are straightforward. First, he picks one of three possible values for x . Then, after Eloise chooses a value for y , he picks a value for z (that may depend on x and y). Thus, a pure strategy for Abelard can be encoded by a pair (a, f) , where $a \in \{1, 2, 3\}$ and $f: \{1, 2, 3\}^2 \rightarrow \{1, 2, 3\}$, while every such pair corresponds to a different pure strategy. It follows that Abelard has $3 \cdot 3^2 = 9$ pure strategies. However, suppose Abelard follows the pure strategy encoded by (a, f) . Then, for all $a' \neq a$, the value of $f(a', b)$ is irrelevant because the history $(\varphi_{\text{MH}}, (x, a'), (y, b))$ will never be played. Thus we may assume that, for all $a', b \in \{1, 2, 3\}$, we have $f(a', b) = f(a, b)$, reducing the number of Abelard's pure strategies to $3 \cdot 3 = 9$.

Next observe that, if Abelard assigns z the same value as x or y , then Eloise can win by choosing the appropriate disjunct. Consequently, Abelard should only follow strategies that lead to histories in which the value of z is distinct from the values of x and y , of which there are only six. Let τ_a and τ^a denote the strategies defined by $\tau_a(\varphi_{\text{MH}}) = (x, a) = \tau^a(\varphi_{\text{MH}})$ and

$$\begin{aligned}\tau_a(\varphi_{\text{MH}}, (x, a), (y, b)) &= (z, \min(\{1, 2, 3\} \setminus \{a, b\})), \\ \tau^a(\varphi_{\text{MH}}, (x, a), (y, b)) &= (z, \max(\{1, 2, 3\} \setminus \{a, b\})).\end{aligned}$$

Eloise has even more strategies than Abelard. Each of Eloise's pure strategies can be encoded by a quadruple (b, g, h, j) , where $b \in \{1, 2, 3\}$ indicates the initial value she assigns to y , the functions

$$\begin{aligned}g: \{1, 2, 3\}^3 &\rightarrow \{(x = z \vee y = z), (\exists y/x) x = y\} \\ h: \{1, 2, 3\}^3 &\rightarrow \{x = z, y = z\}\end{aligned}$$

indicate her choices of disjuncts, and the function

$$j: \{1, 2, 3\}^2 \rightarrow \{1, 2, 3\}$$

indicates the final value she assigns to y . Thus she has a total of $3 \cdot 2^3 \cdot 2^3 \cdot 3^2 = 2^6 \cdot 3^3 = 216$ pure strategies. Fortunately, we only need to consider a small fraction of them. For starters, when faced with the disjunction

$$(x = z \vee y = z) \vee (\exists y/\{x\}) x = y$$

Eloise should only choose the left disjunct if the value of z matches the value of x or y . Otherwise, she should choose the right disjunct. Hence there is only one function worth considering:

$$g(a, b, c) = \begin{cases} (x = z \vee y = z) & \text{if } a = c \text{ or } b = c, \\ (\exists y/\{x\}) x = y & \text{otherwise,} \end{cases}$$

where a is the value assigned to x , b is the value assigned to y , and c is the value assigned to z . Thus, if Eloise ever finds herself faced with the disjunction $(x = z \vee y = z)$, the current assignment will satisfy one of the disjuncts. Therefore, it is sufficient for her to always use the function

$$h(a, b, c) = \begin{cases} x = z & \text{if } a = c, \\ y = z & \text{otherwise.} \end{cases}$$

When faced with the existential subformula $(\exists y/\{x\}) x = y$, Eloise should never set the final value of y equal to the value of z because she knows the values of x and z are distinct. (Were they

the same, she would have chosen the other disjunct.) Similarly, she also knows that the values of y and z will differ. Thus she need only consider functions

$$j: \left(\{1, 2, 3\}^2 \setminus \{(b, b) : b \in \{1, 2, 3\}\} \right) \rightarrow \{1, 2, 3\}$$

such that $j(b, c) \neq c$. There are $2^6 = 64$ such functions, two of which are of particular interest. Let j_{stick} be the function such that for all $b, c \in \{1, 2, 3\}$ we have $j_{\text{stick}}(b, c) = b$, and let j_{switch} be the function such that $j_{\text{switch}}(b, c)$ is the unique element in $\{1, 2, 3\} \setminus \{b, c\}$.

Let σ_b denote the pure strategy for Eloise encoded by the quadruple $(b, g, h, j_{\text{stick}})$, and let σ'_b be the strategy encoded by $(b, g, h, j_{\text{switch}})$. Suppose that σ and τ are rational strategies for Eloise and Abelard, respectively, and that the pair (σ, τ) induce the terminal history

$$\left(\varphi_{\text{MH}}, (x, a), (y, b), (z, c), (\exists y/\{x\}) x = y, (y, d) \right),$$

where $a \neq c \neq b$ and $c \neq d$. Observe that this is the same as the terminal history induced by (σ_b, τ) or (σ'_b, τ) , depending on whether $b = d$ or $b \neq d$, respectively.

Figure 5 highlights the terminal histories that are induced by pairs of rational strategies. By inspection, one can see that these histories form a subtree that is isomorphic to the game tree for the Monty Hall problem shown in Figure 1. Moreover, one can check that two histories are indistinguishable to the contestant if and only if the corresponding histories are indistinguishable to Eloise. Thus, the Monty Hall problem has the same strategic form as the semantic game $G(\mathbf{M}, \varphi_{\text{MH}})$ if we assume that Abelard and Eloise follow rational strategies. Every finite game has a mixed-strategy equilibrium involving only rational strategies [18, Proposition 122.1]. Thus, our previous analysis of the Monty Hall problem shows that the truth value of φ_{MH} is $2/3$.⁸

To conclude this section, let us briefly analyze a variant of the Monty Hall problem in which the host is not required to offer the contestant the opportunity to switch doors. That is, Monty Hall is allowed to open any of the three doors, including the door initially chosen by the contestant or the door containing the prize.

1. If the host opens the contestant's initial door or the door containing the prize, the contestant wins if she guessed correctly and loses if she did not.
2. If the host opens neither the contestant's initial door nor the door containing the prize, he then offers the contestant the opportunity to switch doors.

This scenario can be modeled by the IF sentence φ'_{MH} ,

$$\forall x (\exists y/\{x\}) \forall z \left[(x = y = z) \vee \left[\neg(z = x \neq y) \wedge \neg(z = y \neq x) \wedge (\exists y/\{x\}) x = y \right] \right].$$

A portion of the game tree for $G(\mathbf{M}, \varphi'_{\text{MH}})$ is shown in Figure 6. If the initial values of all three variables are equal, Eloise can win by choosing the disjunct $x = y = z$. In contrast, if Eloise initially assigns y a different value than x , then Abelard can win by setting the value of z equal to the value of x or y . Thus, if x and y are initially assigned the same value, Abelard should pick a different value for z , forcing Eloise to choose the right disjunct, after which Abelard is forced to pick the rightmost conjunct, giving Eloise the opportunity to assign y a new value that may depend on the values of y and z , but not x .

⁸A similar analysis of the Monty Hall problem as the semantic game of an IF sentence appears in Sandu [20, pages 244–245].

and define σ'_b similarly except that

$$\sigma'_b(h_{abc}, \dots, (\exists y/\{x\}) x = y) = (y, b'),$$

where b' is the unique element in $\{1, 2, 3\} \setminus \{b, c\}$ when b and c are distinct, otherwise $b' = \min(\{1, 2, 3\} \setminus \{b\})$. Equivalently, we could have defined $b' = \max(\{1, 2, 3\} \setminus \{b\})$ when $b = c$.

Observe that, if Abelard follows τ_a or τ^a and Eloise follows σ_b or σ'_b , then whenever play reaches the formula $(\exists y/\{x\})x = y$, the values x and y will be the same while the value of z will be different. Hence, she should not change the value of y .

Now define the mixed strategies $\mu, \mu' \in \Delta(S_\exists)$ and $\nu \in \Delta(S_\forall)$ by

$$\nu(\tau_a) = 1/6 = \nu(\tau^a),$$

$$\mu(\sigma_b) = 1/3 = \mu'(\sigma'_b).$$

Then $U_\exists(\mu, \nu) = 1/3$ because Eloise wins if and only if her initial guess is correct. It is easy to verify that $\langle \mu, \nu \rangle$ is a mixed-strategy equilibrium, since Abelard cannot improve his expected utility by favoring certain values of x over others. Nor does it matter which value he assigns to z when the initial values of x and y are the same. For her part, Eloise will win if and only if her initial guess is correct, assuming Abelard follows ν . Thus the value of the semantic game $G(\mathbf{M}, \varphi'_{\text{MH}})$ is $1/3$ (see Figure 7).

$x = 1$		$x = 2$		$x = 3$	
$y = 1$ $z = 2$	$y = 1$ $z = 3$	$y = 1$ $z = 2$		$y = 1$ $z = 3$	
$y = 2$ $z = 1$		$y = 2$ $z = 1$	$y = 2$ $z = 3$	$y = 2$ $z = 3$	
$y = 3$ $z = 1$		$y = 3$ $z = 1$		$y = 3$ $z = 1$	$y = 3$ $z = 2$

Figure 7: A mixed-strategy equilibrium for φ'_{MH}

4 Stochastic IF logic

Another variant of the Monty Hall problem involves treating the host as a disinterested party rather than as a malevolent opponent. If Monty Hall places the prize behind each door with equal probability and opens each door equally often (whether or not it conceals the prize or was chosen by the contestant), then, on those occasions when Monty happens to open a door that neither contains the prize nor was chosen by the contestant, the prize will be found behind each of the remaining doors with equal probability [6, pages 935–936].

To model the scenario just described, we must go beyond ordinary IF logic by adding chance moves to our semantic games. We imagine that such moves are taken by a third player (Nature) who is indifferent to the eventual outcome of the game. Chance moves will be indicated by a new connective \times and quantifier $\mathcal{Z}x$.⁹

Definition. Given any first-order vocabulary, a *stochastic IF formula* is a member of the smallest set $\text{IF}(\mathcal{Z})$ that satisfies the following conditions:

- Every IF formula belongs to $\text{IF}(\mathcal{Z})$.
- If $\varphi, \psi \in \text{IF}(\mathcal{Z})$, then $(\varphi \times \psi) \in \text{IF}(\mathcal{Z})$.
- If $\varphi \in \text{IF}(\mathcal{Z})$, and x is a variable, then $\mathcal{Z}x\varphi \in \text{IF}(\mathcal{Z})$.

An *stochastic IF sentence* is a stochastic IF formula with no free variables.

Definition. Let φ be a stochastic IF sentence, and let \mathbf{M} be a suitable structure. The *semantic game* $G(\mathbf{M}, \varphi)$ is defined as before with the following amendments:

- There are three players, Nature (\mathcal{Z}), Eloise (\exists), and Abelard (\forall).
- If ψ is $\chi_1 \times \chi_2$, then $H_{\chi_i} = \{h \frown \chi_i : h \in H_{\chi_1 \times \chi_2}\}$.
- If ψ is $\mathcal{Z}x\chi$, then $H_{\chi} = \{h \frown (x, a) : h \in H_{\mathcal{Z}x\chi}, a \in M\}$.
- The player function is redefined to make the new connectives and quantifiers moves for Nature:

$$P(h) = \begin{cases} \mathcal{Z} & \text{if } h \in H_{\chi_1 \times \chi_2} \text{ or } h \in H_{\mathcal{Z}x\chi}, \\ \exists & \text{if } h \in H_{\chi_1 \vee_W \chi_2} \text{ or } h \in H_{(\exists x/W)\chi}, \\ \forall & \text{if } h \in H_{\chi_1 \wedge_W \chi_2} \text{ or } h \in H_{(\forall x/W)\chi}. \end{cases}$$

Let $H_{\mathcal{Z}} = P^{-1}(\mathcal{Z})$ be the set of histories in which Nature is the active player.

- Nature’s indistinguishability relation $\sim_{\mathcal{Z}}$ is the identity relation. That is, all of Nature’s information sets are singletons.
- Nature receives no utility regardless of the outcome of the game. For every terminal history $h \in Z$, we have $u_{\mathcal{Z}}(h) = 0$.

⁹Sandu introduces what he calls *probabilistic quantifiers*, denoted μx , where μ is a probability distribution over a finite universe [20, page 246]. We have adapted the backward-S notation from Alexey Radul’s blog post: <http://alexey.radul.name/ideas/2014/stochasticity-is-a-quantifier/>

Since Nature has no reason to prefer one action over another, we cannot reason endogenously about which actions Nature will take. Instead, we will assume that Nature follows a behavioral strategy that is fixed in advance and known to all of the other players. Thus, we will treat Nature's strategy as an exogenous parameter.

In an extensive game with chance moves, a pure-strategy profile for the players other than Nature does not uniquely determine a terminal history; it determines a probability distribution over the set of terminal histories. Thus, although we cannot predict the exact payoffs the players will receive based only on their own actions, we can compute the expected utility for each player given any mixed/behavioral-strategy profile.

Definition. The *expectiminimax value* of a two-player, constant-sum extensive game with chance moves, relative to a fixed behavioral strategy λ for Nature, is

$$\min_{\nu \in \Delta(S_{II})} \max_{\mu \in \Delta(S_I)} U_I(\lambda, \mu, \nu) = U_I(\lambda, \mu^*, \nu^*) = \max_{\mu \in \Delta(S_I)} \min_{\nu \in \Delta(S_{II})} U_I(\lambda, \mu, \nu).$$

Definition. Let φ be a stochastic IF sentence, let \mathbf{M} be a suitable finite structure, and let λ be a behavioral strategy for Nature in the semantic game $G(\mathbf{M}, \varphi)$. The *truth value* of φ in \mathbf{M} (relative to λ) is the expectiminimax value of $G(\mathbf{M}, \varphi)$ when Nature follows λ . We will use the notation $\mathbf{M} \models^v \varphi(\lambda)$ to express the fact that v is the truth value of φ in \mathbf{M} (relative to λ).

Example. Consider the following stochastic generalization of the Matching Pennies sentence:¹⁰

$$\forall x (\exists y / \{x\}) \mathcal{C}z [x = y = z]$$

When played on the two-element structure $\mathbf{2} = \{0, 1\}$, the semantic game for the above sentence models the scenario in which two players each turn a coin to Heads or Tails, but now there is a third coin that Nature tosses in secret. The first player (Eloise) wins if all three coins match; otherwise she loses.

If we assume that Nature's coin is biased, so that Nature follows the behavioral strategy defined by $\lambda(0) = 1/3$ and $\lambda(1) = 2/3$, then the semantic game can be represented by the game tree shown in Figure 8, where Abelard follows the mixed strategy defined by $\nu_q(0) = q$ and $\nu_q(1) = 1 - q$, while Eloise follows the mixed strategy defined by $\mu_p(0) = p$ and $\mu_p(1) = 1 - p$. Then Eloise's expected utility is $U_{\exists}(\lambda, \mu_p, \nu_q) = \frac{1}{3}pq + \frac{2}{3}(1-p)(1-q)$. Using the second-derivative test, we can show that there is an equilibrium when $p = 2/3$ and $q = 2/3$. Hence the truth value of the sentence is $U_{\exists}(\lambda, \mu_{2/3}, \nu_{2/3}) = 2/9$. This equilibrium is depicted in Figure 9. The reader should imagine that the square on the left labeled $z = 0$ has a vertical "thickness" of $1/3$, while the square on the right labeled $z = 1$ has a thickness of $2/3$, which accounts for the fact that Nature's coin is biased.

¹⁰Sandu considers a different stochastic variant of the Matching Pennies sentence [20, page 246].

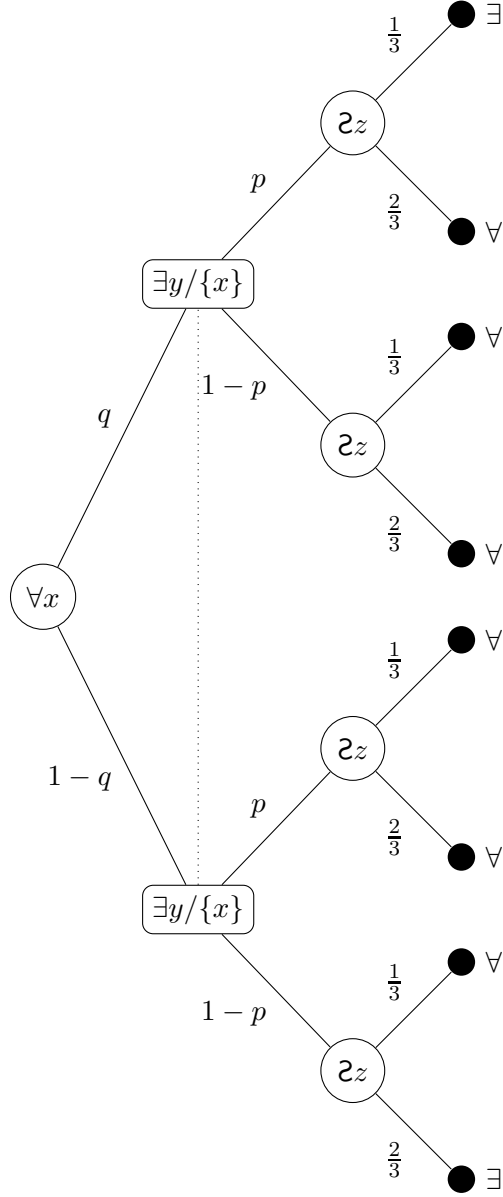


Figure 8: The semantic game for the stochastic matching pennies sentence

$z = 0$		$z = 1$	
$x = 0$ $y = 0$	$x = 1$ $y = 0$	$x = 0$ $y = 0$	$x = 1$ $y = 0$
$x = 0$ $y = 1$	$x = 1$ $y = 1$	$x = 0$ $y = 1$	$x = 1$ $y = 1$

Figure 9: An equilibrium for the stochastic Matching Pennies sentence

We are now ready to model the variant of the Monty Hall problem in which the host is indifferent to the outcome as the semantic game of a stochastic IF sentence. We will assume that the contestant wins if Monty Hall reveals that the door she initially chose contains the prize and loses if he reveals that her door does not contain the prize or that some other door does contain the prize. Otherwise, she is offered the opportunity to switch doors. We will further assume that the prize is placed behind each door with probability $1/3$, and that Monty Hall opens each door with probability $1/3$.

We will denote the following stochastic variant of φ'_{MH} by $\varphi'_{\text{MH}\mathcal{Z}}$,

$$\mathcal{Z}x(\exists y/\{x\})\mathcal{Z}z \left[(x = y = z) \vee \left[\neg(z = x \neq y) \wedge \neg(z = y \neq x) \wedge (\exists y/\{x\}) x = y \right] \right],$$

and consider the semantic game $G(\{1, 2, 3\}, \varphi'_{\text{MH}\mathcal{Z}})$ in which Nature follows a behavioral strategy λ that assigns values to x and z according to a uniform probability distribution.

Although he no longer assigns values to x and z , Abelard still plays a role in the semantic game because of the subformula

$$\neg(x \neq y = z) \wedge \neg(x = z \neq y) \wedge (\exists y/\{x\}) x = y,$$

which we treat as a ternary conjunction (denoted ψ below). Let $h_{abc} = (\varphi'_{\text{MH}\mathcal{Z}}, (x, a), (y, b), (z, c))$, let τ be the pure strategy for Abelard defined by

$$\tau(h_{abc} \frown \psi) = \begin{cases} \neg(z = x \neq y) & \text{if } c = a \neq b, \\ \neg(z = y \neq x) & \text{if } c = b \neq a, \\ (\exists y/\{x\}) x = y & \text{otherwise,} \end{cases}$$

and let ν be the mixed strategy for Abelard such that $\nu(\tau) = 1$.

After the initial values of x , y , and z have been set, Eloise should choose the left disjunct if and only if all three values are the same. Let $h_a = (\varphi'_{\text{MH}\mathcal{Z}}, (x, a))$, and define $\sigma_b \in S_{\exists}$ by

$$\sigma_b(h_a) = (y, b),$$

$$\sigma_b(h_{abc}) = \begin{cases} x = y = z & \text{if } a = b = c, \\ \psi & \text{otherwise,} \end{cases}$$

$$\sigma_b((h_{abc} \frown \psi) \frown (\exists y/\{x\}) x = y) = \min(\{1, 2, 3\} \setminus \{c\}).$$

Define $\sigma^b \in S_{\exists}$ similarly except that $\sigma^b((h_{abc} \frown \psi) \frown (\exists y/\{x\}) x = y) = \max(\{1, 2, 3\} \setminus \{c\})$. Let $\mu \in \Delta(S_{\exists})$ be the mixed strategy according to which $\mu(\sigma_b) = 1/6 = \mu(\sigma^b)$. (Strictly speaking, the formulas $\neg(x \neq y = z)$ and $\neg(x = z \neq y)$ abbreviate disjunctions, but we will treat them as literals.)

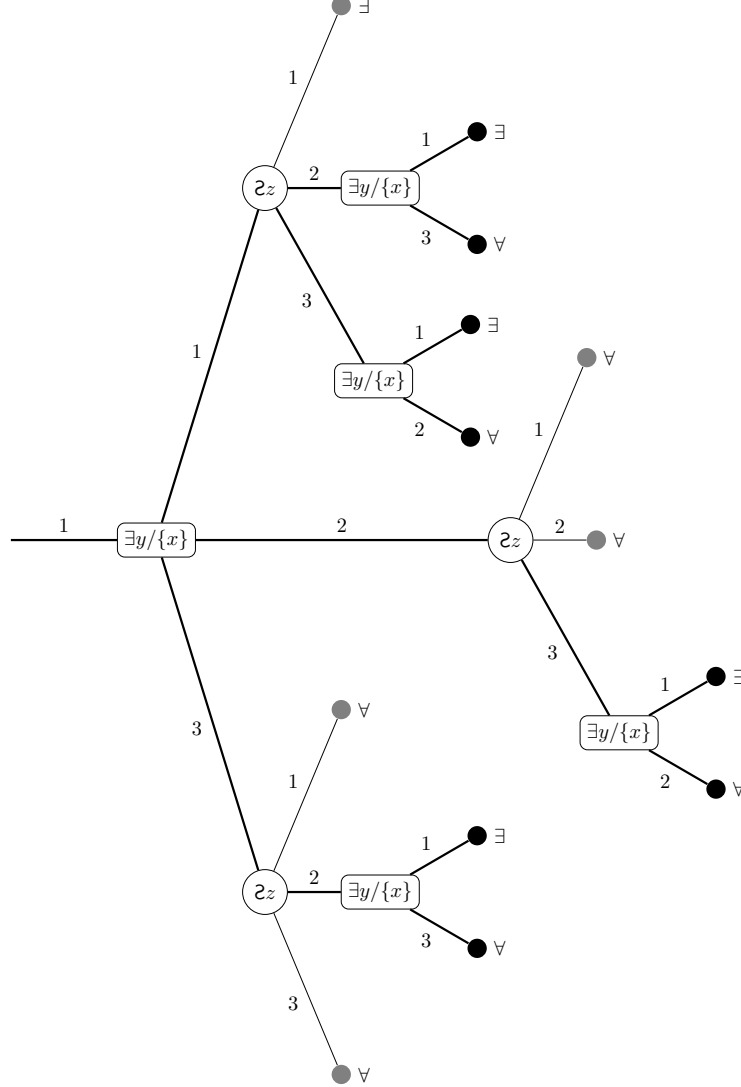


Figure 10: A portion of the semantic game for φ'_{MHS}

Figure 10 shows the histories that occur when $x = 1$, Abelard follows ν , and Eloise follows μ . If initially $y = 1$ and $z = 1$, Eloise wins by choosing the left disjunct $x = y = z$. If instead $z = 2$ or $z = 3$, then Eloise will choose the right disjunct ψ , after which Abelard will pick the right conjunct. Eloise will then reset the value of y to be distinct from the value of z , winning when she sticks with her initial choice. If initially $y = 2$ or $y = 3$, then Eloise will choose ψ , after which Abelard will win, if possible, by picking the appropriate conjunct based on the value of z . Otherwise, he will pick $(\exists y/\{x\}) x = y$, in which case Eloise will reset the value of y to be distinct from the value of z , but this time she will lose when she sticks with her original choice and win when she switches.

If we restrict our attention to those histories, highlighted in Figure 10, in which Nature selects

a value of z that differs from the values of both x and y , we can see that Eloise wins one-half of the time when she sticks with her original y -value, and wins one-half of the time when she switches.

To conclude this section, let us consider the following stochastic variant φ_{MH2} of the original Monty Hall sentence,¹¹

$$\mathcal{Z}x(\exists y/\{x\})\mathcal{Z}z\left[(z \neq x \wedge z \neq y) \implies (\exists y/\{x\})x = y\right],$$

which is an abbreviation for

$$\mathcal{Z}x(\exists y/\{x\})\mathcal{Z}z\left[(z = x \vee z = y) \vee (\exists y/\{x\})x = y\right].$$

Once again we assume that Nature follows a behavioral strategy λ that assigns values to x and z according to a uniform probability distribution. Since the above sentence has no conjunctions or universal quantifiers, Abelard plays no role in its semantic game. Let $\sigma_b, \sigma'_b \in S_{\exists}$ be defined as before, with Eloise initially setting the value of y to be b , then sticking or switching, respectively. Let $\mu \in \Delta(S_{\exists})$ be the mixed strategy according to which $\mu(\sigma_b) = 1/6 = \mu(\sigma'_b)$.

Unlike Abelard, Nature has no qualms about setting the value of z equals to the value of x or y . In fact, if Nature follows λ and Eloise follows μ , then Eloise will win by choosing the disjunct $(z = x \vee z = y)$ in 5/9 of all plays. In the remaining plays, exactly half will satisfy $x = y$. Thus Eloise will win an additional 2/9 of all plays by correctly guessing the value of x , regardless of whether she sticks with her initial guess or switches. Thus, the truth value of φ_{MH2} is 7/9.

¹¹Sandu considers a variant of φ_{MH} in which only the first quantifier is stochastic [20, page 247].

5 Sleeping Beauty

We now turn our attention to a related puzzle, called the Sleeping Beauty problem, popularized by Adam Elga:

Some researchers are going to put you to sleep. During the two days that your sleep will last, they will briefly wake you up either once or twice, depending on the toss of a fair coin (Heads: once; Tails: twice). After each waking, they will put you back to sleep with a drug that makes you forget that waking.

When you are first awakened, to what degree ought you believe that the outcome of the coin toss is Heads? [5, page 143]

Two answers, $1/2$ and $1/3$, have been defended in the literature. The proponents of each answer are known as halfers and thirders, respectively.

Elga argues in favor of $1/3$. Suppose the first waking occurs on Monday, the second on Tuesday. Then, immediately after waking, you will be in one of three possible situations: H_1 , the coin landed Heads and it is Monday; T_1 , the coin landed Tails and it is Monday; or T_2 , the coin landed Tails and it is Tuesday. According to Elga, all three are equally likely. For suppose that after waking you are told that the coin landed Tails. At that moment, you know that you are in situation T_1 or T_2 , but—because of the drug—you cannot tell which. Furthermore, you have no reason to suspect that T_1 is more or less likely than T_2 , given that the coin landed tails. Consequently,

$$P(T_1 | T_1 \cup T_2) = P(T_2 | T_1 \cup T_2),$$

which implies $P(T_1) = P(T_2)$ [5, page 144].

Now suppose instead that, after waking you up, the researchers inform you that it is Monday. Then you would know that you are in situation H_1 or T_1 , which of the two being determined by the toss of a fair coin. In fact, it doesn't matter whether the researchers toss the coin before or after waking you the first time. (You are woken on Monday regardless of the outcome.) So you might suppose that the researchers have yet to toss the coin, in which case $P(H_1 | H_1 \cup T_1)$ is simply the probability that a fair coin that has yet to be tossed will land Heads. Hence $P(H_1 | H_1 \cup T_1) = 1/2$. A simple calculation then shows that $P(H_1) = P(T_1)$. Therefore

$$P(H_1) = P(T_1) = P(T_2),$$

which implies $P(H_1) = 1/3$ [5, page 145].

David Lewis argues, contra Elga, that you do not gain any information relevant to the outcome of a fair coin toss simply by being awake since you knew all along that you would be awakened at least once during the experiment. Lewis agrees with Elga that $P(T_1) = P(T_2)$, but challenges his assertion that $P(H_1 | H_1 \cup T_1) = 1/2$. Instead, Lewis claims that one should assign equal prior probabilities to Heads and Tails:

$$P(H_1) = 1/2 = P(T_1 \cup T_2).$$

It follows that $P(T_1) = 1/4 = P(T_2)$. Consequently, when the researchers inform you that it is Monday, you should update your beliefs accordingly:

$$P(H_1 | H_1 \cup T_1) = \frac{P(H_1)}{P(H_1) + P(T_1)} = 2/3.$$

Lewis points out that he and Elga both agree that probability of Heads depends on what day it is, since $P(H_1 | H_1 \cup T_1) = P(H_1) + 1/6$, but they disagree as to whether the prior or the posterior probability should be taken to be $1/2$ [13].

Cian Dorr comes to Elga's defense by considering a variation of the experiment in which the researchers have two amnesia-inducing drugs instead of just one [4]. If the coin lands Tails, they will administer the same drug as before, so that your experience is exactly the same as in the original experiment. If the coin lands Heads, however, they will administer a weaker drug so that, when you wake up the second time, your memories of the first awakening will be delayed for exactly one minute. Thus, immediately after being awoken, it is possible that you are in the situation H_2 , the coin landed heads and it is Tuesday. Since your subjective experience in each of the four situations is identical, you should consider them all to be equally likely:

$$P(H_1) = P(H_2) = P(T_1) = P(T_2) = 1/4.$$

After one minute passes, either you will remember being woken up on Monday, or you will not. If you do, your memories will confirm that you are in the situation H_2 . If you do not, your failure to remember (drug-induced or not) is evidence that you are currently experiencing H_1 , T_1 , or T_2 , at which point the probability that the coin landed Heads is

$$P(H_1 \mid H_1 \cup T_1 \cup T_2) = \frac{P(H_1)}{P(H_1) + P(T_1) + P(T_2)} = 1/3.$$

Arntzenius independently developed a similar variant of the Sleeping Beauty problem in which the subject of the experiment is a vivid dreamer who can distinguish wake from dream by pinching herself [1, pages 363–364]. It is worth noting that whether Dorr's (and presumably Arntzenius's) variations are analogous to the original version of the problem is controversial [3, 12].

We are now ready to formalize the Sleeping Beauty problem using stochastic IF logic. At the moment she is awakened, Beauty thinks to herself: "Given that I am awake, a fair coin must have been tossed, but I don't know whether it landed Heads or Tails. Furthermore, because of the amnesia-inducing drug I may have been given, I am unsure whether it is Monday or Tuesday." To help determine what her credence should be, she decides to model her predicament using the following stochastic IF sentence φ_{SB} ,

$$\mathcal{Z}x\mathcal{Z}t \left[\text{Awake}(x, t) \implies (\text{Heads}(x) \vee_{/\{x, t\}} \text{Tails}(x)) \right],$$

interpreted in a structure \mathbf{M} with universe $\{1, 2\}$ and equipped with the following relations:

$$\begin{aligned} \text{Heads}^{\mathbf{M}} &= \{1\} = \text{Monday}^{\mathbf{M}} \\ \text{Tails}^{\mathbf{M}} &= \{2\} = \text{Tuesday}^{\mathbf{M}} \\ \text{Awake}^{\mathbf{M}} &= \{(1, 1), (2, 1), (2, 2)\}. \end{aligned}$$

Here x represents the result of the coin toss, and t represents the current time.

Next Beauty analyzes the semantic game $G(\mathbf{M}, \varphi_{\text{SB}})$, which begins with Nature selecting values for x and t . Recall that the implication inside the square brackets is an abbreviation for

$$\neg \text{Awake}(x, t) \vee (\text{Heads}(x) \vee_{/\{x, t\}} \text{Tails}(x)).$$

Thus, Eloise must first choose between $\neg \text{Awake}(x, t)$ and $(\text{Heads}(x) \vee_{/\{x, t\}} \text{Tails}(x))$. If she chooses the left disjunct, the game ends. If she chooses the right disjunct, she must choose between the atomic formulas $\text{Heads}(x)$ or $\text{Tails}(x)$ without knowing the value of x or t .

Considering the players' possible strategies, Beauty reasons that Nature should follow a behavioral strategy λ according to which $\lambda(x := 1) = 1/2 = \lambda(x := 2)$ since the coin is fair. Beauty further assumes that $\lambda(t := 1) = 1/2 = \lambda(t := 2)$ since there is no reason for Nature to prefer one day over the other.

Assuming λ is fixed, how should Eloise play? After the partial history $h_{ab} = (\varphi_{\text{SB}}, (x, a), (t, b))$, she is confronted with the unslashed disjunction

$$\neg \text{Awake}(x, t) \vee (\text{Heads}(x) \vee_{/\{x, t\}} \text{Tails}(x)).$$

Eloise should clearly choose the left disjunct if and only if $(a, b) = (1, 2)$, so let

$$\sigma_1(h_{ab}) = \begin{cases} \neg \text{Awake}(x, t) & \text{if } (a, b) = (1, 2), \\ (\text{Heads}(x) \vee_{/\{x, t\}} \text{Tails}(x)) & \text{otherwise,} \end{cases}$$

$$\sigma_1(h_{ab} \frown (\text{Heads}(x) \vee_{/\{x, t\}} \text{Tails}(x))) = \text{Heads}(x),$$

and define σ_2 similarly except that $\sigma_2(h_{ab} \frown (\text{Heads}(x) \vee_{/\{x, t\}} \text{Tails}(x))) = \text{Tails}(x)$. All of her other strategies are dominated, so let μ_p be the mixed strategy over $\{\sigma_1, \sigma_2\}$ such that $\mu_p(\sigma_1) = p$ and $\mu_p(\sigma_2) = 1 - p$.

As shown in Figure 11, when $(a, b) = (1, 1)$, Eloise wins if she follows σ_1 and loses if she follows σ_2 . If $(a, b) = (1, 2)$, Eloise will win regardless of which strategy she follows. If $a = 2$, Eloise wins if she follows σ_2 and loses if she follows σ_1 . Hence Eloise's expected utility from μ_p is

$$U(\lambda, \mu_p) = \frac{p}{4} + \frac{1}{4} + \frac{1-p}{2} = \frac{3-p}{4}.$$

Thus, Eloise wins exactly half of the plays in which she follows σ_1 and three-quarters of the plays in which she follows σ_2 . Hence her optimal strategy is to always follow σ_2 . Therefore $\mathbf{M} \models^{3/4} \varphi_{\text{SB}}(\lambda)$.

However, Beauty is not so much interested in the truth value of the sentence φ_{SB} relative to λ as she is in Eloise's chances of winning when $(a, b) \in \text{Awake}^{\mathbf{M}}$. If we exclude those histories in which $(a, b) = (1, 2)$ (grayed out in Figure 11), then Eloise's chances of winning become

$$\frac{\frac{p}{4} + \frac{1-p}{2}}{\frac{3}{4}} = \frac{2-p}{3}.$$

Thus, Eloise wins one-third of the histories in which she chooses $\text{Heads}(x)$ and two-thirds of the histories in which she chooses $\text{Tails}(x)$. Therefore, Beauty concludes that the probability that the coin landed Heads is $1/3$.¹²

A halfer might object that Beauty made a mistake by considering the wrong stochastic IF sentence. Instead of φ_{SB} , Beauty should have considered the following alternative sentence φ'_{SB} ,

$$\exists x \forall t [\text{Awake}(x, t) \implies (\text{Heads}(x) \vee_{/\{x, t\}} \text{Tails}(x))],$$

in which the universal player picks the value of t instead of Nature. If she had, she would have found that Nature's behavioral strategy $\lambda(x := 1) = 1/2 = \lambda(x := 2)$ is stipulated by the fact that the coin is fair, and that Eloise's pure and mixed strategies are the same as before. Since Eloise

¹²Beauty's reasoning most closely matches Horgan's argument [12] that one should assign each of the four hypotheses H_1 , H_2 , T_1 , and T_2 a prior probability of $1/4$. See also Hitchcock's diachronic Dutch Book argument [11].

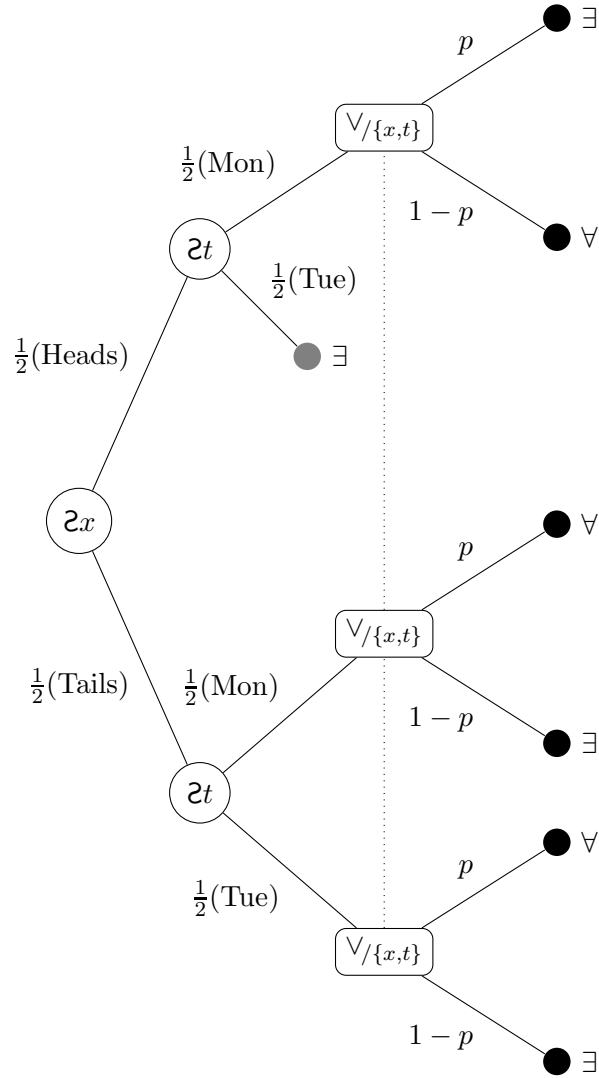


Figure 11: The semantic game for φ_{SB}

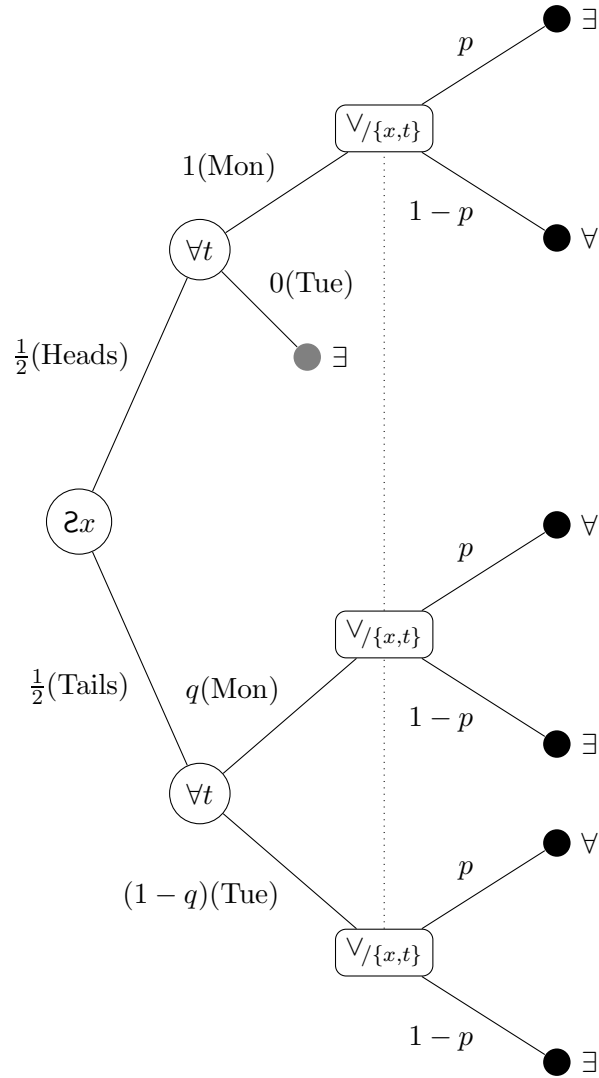


Figure 12: The semantic game for φ'_{SB}

wins any history in which $(a, b) = (1, 2)$, Abelard should always set $t := 1$ when $x := 1$. Let ν_q be the mixed strategy for Abelard such that

$$\begin{aligned}\nu_q(x := 1, t := 1) &= 1 & \text{and} & & \nu_q(x := 1, t := 2) &= 0, \\ \nu_q(x := 2, t := 1) &= q & \text{and} & & \nu_q(x := 2, t := 2) &= 1 - q.\end{aligned}$$

Eloise's expected utility when she follows μ_p , and Abelard follows ν_q is

$$U(\lambda, \mu_p, \nu_q) = \frac{p}{2} + \frac{(1-p)q}{2} + \frac{(1-p)(1-q)}{2} = 1/2.$$

Notice that $U(\lambda, \mu_p, \nu_q)$ depends on neither p nor q . Thus, Eloise wins exactly half of the time no matter how she guesses, reflecting the fact that the coin is fair. Furthermore, if we restrict our attention to those histories in which the value of t is 1 (i.e., Monday), Eloise's chances of winning become

$$\frac{\frac{1}{2}p + \frac{1}{2}(1-p)q}{\frac{1}{2} + \frac{1}{2}q} = \frac{p + (1-p)q}{1 + q}$$

When $q = 0$, Eloise wins with probability p ; when $q = 1/2$, she wins with probability $(p+1)/3$; and when $q = 1$, she wins with probability $1/2$. This shows that, if the researchers always wake Beauty on Monday when the coin lands Heads, and always wake her on Tuesday (and only Tuesday) when the coin lands Tails, then, immediately after waking, her credence that the coin landed Heads should be $1/2$, but, after being told that it is Monday, she becomes certain that the coin landed Heads. At the other extreme, if the researchers always wake Beauty on Monday (and never on Tuesday), regardless of the result of the coin toss, then learning that it is Monday gives her no information about the coin toss, so her credence should remain $1/2$. Finally, suppose the researchers wake Beauty exactly once during the experiment: on Monday if the coin lands Heads, and on Monday *or* Tuesday (with equal probability) if the coin lands Tails. Then her initial credence that the coin landed Heads should be $1/2$, but should increase to $2/3$ after she learns that it is Monday.

In summary, when Abelard follows $\nu_{1/2}$ and Eloise follows σ_1 , she will win one-half of all plays, and two-thirds of the plays in which the value of t is 1. This is the correct result according to halfers. Unfortunately for them, Abelard's strategy $\nu_{1/2}$ is dominated by ν_1 , so it would be irrational for him to follow it.

The preceding analysis suggests another possibility, however. Instead of proposing the alternative stochastic IF sentence φ'_{SB} , a halfer might assert that, in the semantic game for φ_{SB} , Nature should follow the alternative behavioral strategy λ' according to which

$$\begin{aligned}\lambda'(x := 1, t := 1) &= 1/2 & \text{and} & & \lambda'(x := 1, t := 2) &= 0, \\ \lambda'(x := 2, t := 1) &= 1/4 & \text{and} & & \lambda'(x := 2, t := 2) &= 1/4.\end{aligned}$$

However, Beauty's prior analysis of φ'_{SB} shows that the game depicted in Figure 13 does not accurately formalize her predicament. Rather, it formalizes an experiment in which the subject is awakened exactly once: on Monday if the coin lands Heads, and on Monday *or* Tuesday (with equal probability) if the coin lands Tails. Thus, we are forced to conclude that Lewis' argument in favor of $1/2$ is correct for this variant of the Sleeping Beauty problem, but is incorrect when applied to the original version.

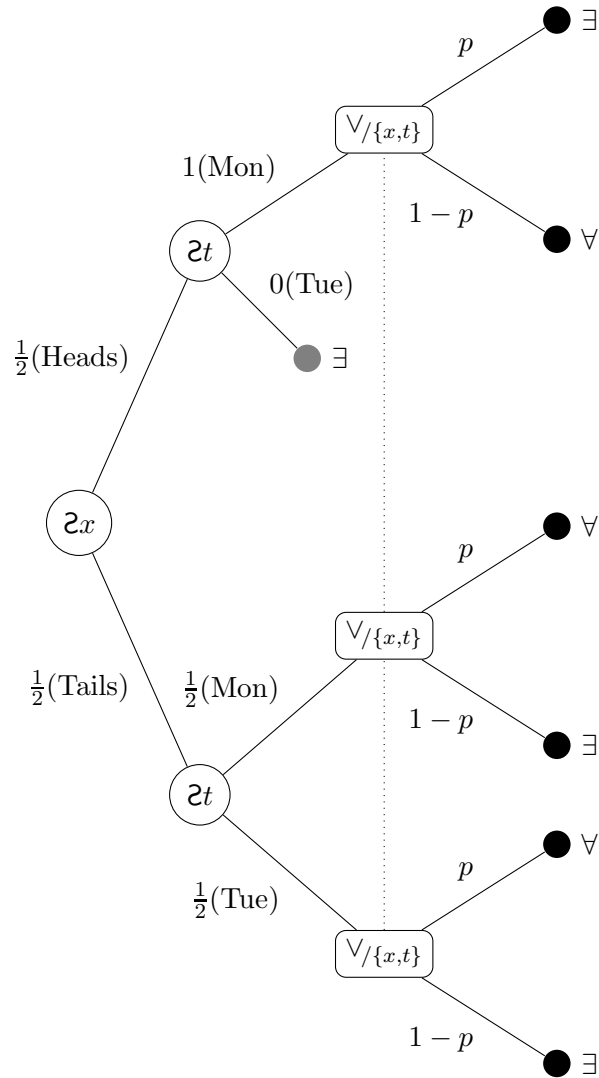


Figure 13: The semantic game for φ_{SB} when Nature follows λ'

6 Conclusion

The dual goals of the present article have been to deepen our understanding of the Monty Hall and Sleeping Beauty problems by formalizing them in a logical language, while simultaneously using these famous problems to introduce a natural extension of first-order logic with imperfect information.

We began by showing how the original Monty Hall problem and a variant in which the host is not required to offer the contestant the opportunity to switch doors could be viewed as semantic games of IF sentences, where Abelard plays the role of the host and Eloise plays the role of the contestant. We then modeled two further variants of the Monty Hall problem in which the host is indifferent to the outcome as semantic games with chance moves.

We also showed how Beauty could use semantic games with chance moves to analyze her predicament, leading to the conclusion that the thirders are correct. Finally, we explained how semantic games with chance moves could be used to demonstrate that Lewis's argument in favor of $1/2$ applies to a variant of the Sleeping Beauty problem. Although we doubt ours will be the last word on the matter, we hope that future contributors to the Sleeping Beauty debate will follow our example by presenting explicitly their formalizations of the problem.

References

- [1] F. Arntzenius. Some problems for conditionalization and reflection. *Journal of Philosophy*, 100:356–370, 2003.
- [2] A. Blass and Y. Gurevich. Henkin quantifiers and complete problems. *Annals of Pure and Applied Logic*, 32:1–16, 1986.
- [3] D. Bradley. Sleeping Beauty: a note on Dorr’s argument for $1/3$. *Analysis*, 63:266–268, 2003.
- [4] C. Dorr. Sleeping Beauty: in defense of Elga. *Analysis*, 62:292–96, 2002.
- [5] A. Elga. Self-locating belief and the Sleeping Beauty problem. *Analysis*, 60:143–47, 2000.
- [6] D. Friedman. Monty Hall’s three doors: Construction and deconstruction of a choice anomaly. *American Economic Review*, 88:933–946, 1998.
- [7] D. Gale and F. M. Stewart. Infinite games with perfect information. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games II*, volume 28 of *Annals of Mathematics Studies*, chapter 13, pages 245–266. Princeton University Press, Princeton, NJ, 1953.
- [8] P. Galliani. Game values and equilibria for undetermined sentences of dependence logic. MSc thesis, ILLC Publications, MoL-2008-08, 2008.
- [9] J. Hintikka. *The Principles of Mathematics Revisited*. Cambridge University Press, Cambridge, 1996.
- [10] J. Hintikka and G. Sandu. Informational independence as a semantical phenomenon. In J. E. Fenstad, I. T. Frolov, and R. Hilpinen, editors, *Logic, Methodology and Philosophy of Science VIII*, volume 126 of *Studies in Logic and the Foundations of Mathematics*, pages 571–589. North-Holland, Amsterdam, 1989.
- [11] C. Hitchcock. Beauty and the bets. *Synthese*, 139:405–420, 2004.
- [12] T. Horgan. Sleeping Beauty awakened: new odds at the dawn of the new day. *Analysis*, 64:10–21, 2004.
- [13] D. Lewis. Sleeping Beauty: reply to Elga. *Analysis*, 61:171–76, 2001.
- [14] A. L. Mann, G. Sandu, and M. Sevenster. *Independence-Friendly Logic: A Game-Theoretic Approach*. Number 386 in London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge, 2011.
- [15] M. Maschler, E. Solan, and S. Zamir. *Game Theory*. Cambridge University Press, New York, 2013.
- [16] J. F. Nash. Equilibrium points in n -person games. *Proceedings of the National Academy of Sciences of the United States of America*, 36:48–49, 1950.
- [17] J. F. Nash. Non-cooperative games. *The Annals of Mathematics*, 54:286–295, 1951.
- [18] M. J. Osborne. *An Introduction to Game Theory*. Oxford University Press, New York, 2004.
- [19] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, Cambridge, Massachusetts, 1994.

- [20] G. Sandu. Languages for imperfect information. In J. van Benthem et al., editors, *Models of Strategic Reasoning*, volume 8972 of *LNCS*, pages 202–251. Springer, Berlin, 2015.
- [21] M. Sevenster and G. Sandu. Equilibrium semantics of languages of imperfect information. *Annals of Pure and Applied Logic*, 161:618–631, 2010.
- [22] J. Tierney. And Behind Door No. 1, a Fatal Flaw. *New York Times*, April 8, 2008.
- [23] J. Tierney. Behind Monty Hall’s Doors: Puzzle, Debate and Answer? *New York Times*, July 21, 1991.
- [24] J. Väänänen. *Dependence Logic: A New Approach to Independence Friendly Logic*. Number 70 in London Mathematical Society Student Texts. Cambridge University Press, Cambridge, 2007.
- [25] J. von Neumann. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, 100:295–320, 1928.
- [26] M. vos Savant. Ask Marilyn. *Parade*, September 8, 1990; December 2, 1990; February 17, 1991.